

IN THE CLAIMS:

The following is a complete listing of the claims in this application, reflects all changes currently being made to the claims, and replaces all earlier versions and all earlier listings of the claims:

1. (Currently Amended): Image processing apparatus, comprising:
  - an image data receiver for receiving image data recorded by a plurality of cameras showing the movements of a plurality of people;
  - a speaker identifier for determining which of the people is speaking;
  - a speech recipient identifier for determining at whom the speaker is looking;
  - a position calculator for determining the position of the speaker and the position of the person at whom the speaker is looking; and
  - [[a]] camera ~~selector~~ selection means for selecting image data from the received image data on the basis of the determined positions of the speaker and the person at whom the speaker is looking, said camera selection means being arranged to select image data in which both the speaker and the person at whom the speaker is looking appear, and  
wherein the camera selection means is arranged to generate quality values representing a quality of the views that at least some of the cameras have of the speaker and the person at whom the speaker is looking, and to select the image data on the basis of which camera has the quality value representing the highest quality.

2.-3. (Canceled).

4. (Currently Amended): Apparatus according to claim [[3]] 1, wherein the camera ~~selector~~ selection means is arranged to determine which of the cameras have a view of the speaker and the person at whom the speaker is looking, and to generate a respective quality value for each camera which has a view of the speaker and the person at whom the speaker is looking.

5. (Currently Amended): Apparatus according to claim [[3]] 1, wherein the camera ~~selector~~ selection means is arranged to generate each quality value in dependence upon the position and orientation of the head of the speaker and the position and orientation of the head of the person at whom the speaker is looking.

6. (Currently Amended): Apparatus according to claim 1, wherein the camera ~~selector~~ selection means comprises:

a data store for storing data defining a camera from which image data is to be selected for respective pairs of positions; and

an image data selector arranged to use data stored in the data store to select the image data in dependence upon the positions of the speaker and the person at whom the speaker is looking.

7. (Original): Apparatus according to claim 1, wherein the speech recipient identifier and the position calculator comprise an image processor for processing the image data from at least one of the cameras to determine at whom the speaker is looking and the positions.

8. (Original): Apparatus according to claim 7, wherein the image processor is arranged to determine the position of each person and at whom each person is looking by processing the image data from the at least one camera.

9. (Original): Apparatus according to claim 7, wherein the image processor is arranged to track the position and orientation of each person's head in three dimensions.

10. (Original): Apparatus according to claim 1, wherein the speaker identifier is arranged to receive speech data from a plurality of microphones each of which is allocated to a respective one of the people, and to determine which of the people is speaking on the basis of the microphone from which the speech data was received.

11. (Original): Apparatus according to claim 1, further comprising a sound processor for processing sound data defining words spoken by the people to generate text data therefrom in dependence upon the result of the processing performed by the speaker

identifier.

12. (Original): Apparatus according to claim 11, wherein the sound processor has associated therewith a store for storing respective voice recognition parameters for each of the people, and a parameter selector for selecting the voice recognition parameters to be used to process the sound data in dependence upon the person determined to be speaking by the speaker identifier.

13. (Original): Apparatus according to claim 11, further comprising a database for storing at least some of the received image data, the sound data, the text data produced by the sound processor and viewing data defining at whom at least the person who is speaking is looking, the database being arranged to store the data such that corresponding text data and viewing data are associated with each other and with the corresponding image data and sound data.

14. (Original): Apparatus according to claim 13, further comprising a data compressor for compressing the image data and the sound data for storage in the database.

15. (Original): Apparatus according to claim 14, wherein the data compressor comprises an encoder for encoding the image data and the sound data as MPEG data.

16. (Original): Apparatus according to claim 13, further comprising a gaze time data generator for generating gaze time data defining, for a predetermined period, the proportion of time spent by a given person looking at each of the other people during the

predetermined period, and wherein the database is arranged to store the gaze time data so that it is associated with the corresponding image data, sound data, text data and viewing data.

17. (Original): Apparatus according to claim 16, wherein the predetermined period comprises a period during which the given person was talking.

18. (Canceled).

19. (Currently Amended): A method of processing image data recorded by a plurality of cameras showing the movements of a plurality of people to select image data for storage, the method comprising:

a speaker identification step of determining which of the people is speaking;

a step of determining at whom the speaker is looking;

a step of determining the position of the speaker and the position of the person at whom the speaker is looking; and

a camera selection step ~~of~~ for selecting image data on the basis of the determined positions of the speaker and the person at whom the speaker is looking, wherein, in the camera selection step, image data is selected in which both the speaker and the person at whom the speaker is looking appear, quality values are generated representing a quality of the views that at least some of the cameras have of the speaker and the person at whom the speaker is looking, and the image data is selected on the basis of which camera has the quality value representing the highest quality.

20.-21. (Canceled).

22. (Currently Amended): A method according to claim ~~[[21]]~~ 19, wherein, in the camera selection step, processing is performed to determine which of the cameras have a view of the speaker and the person at whom the speaker is looking, and to generate a respective quality value for each camera which has a view of the speaker and the person at whom the speaker is looking.

23. (Currently Amended): A method according to claim ~~[[21]]~~ 19, wherein, in the camera selection step, each quality value is generated in dependence upon the position and orientation of the head of the speaker and the position and orientation of the head of the person at whom the speaker is looking.

24. (Original): A method according to claim 19, wherein, in the camera selection step pre-stored data defining a camera from which image data is to be selected for respective pairs of positions is used to select the image data in dependence upon the positions of the speaker and the person at whom the speaker is looking.

25. (Original): A method according to claim 19, wherein, in the steps of determining at whom the speaker is looking and determining the positions of the speaker and the person at whom the speaker is looking, image data from at least one of the cameras is processed to determine at whom the speaker is looking and the positions.

26. (Currently Amended): A method according to claim 25, wherein[[,]] the image data from that at least one camera is processed to determine the position of each person and at whom each person is looking.

27. (Original): A method according to claim 25, wherein image data is processed to track the position and orientation of each person's head in three dimensions.

28. (Original): A method according to claim 19, wherein speech data is received from a plurality of microphones each of which is allocated to a respective one of the people, and, in the speaker identification step, it is determined which of the people is speaking on the basis of the microphone from which the speech data was received.

29. (Original): A method according to claim 19, further comprising a sound processing step of processing sound data defining words spoken by the people to generate text data therefrom in dependence upon the result of the processing performed in the speaker identification step.

30. (Original): A method according to claim 29, wherein the sound processing step includes selecting, from among stored respective voice recognition parameters for each of the people, the voice recognition parameters to be used to process the sound data in dependence upon the person determined to be speaking in the speaker identification step.

31. (Original): A method according to claim 29, further comprising the step of storing in a database at least some of the received image data, the sound data, the text data produced in the sound processing step and viewing data defining at whom at least the person who is speaking is looking, the data being stored in the database such that corresponding text data and viewing data are associated with each other and with the corresponding image data and sound data.

32. (Original): A method according to claim 31, wherein the image data and the sound data are stored in the database in compressed form.

33. (Original): A method according to claim 32, wherein the image data and the sound data are stored as MPEG data.

34. (Original): A method according to claim 31, further comprising the steps of generating data defining, for a predetermined period, the proportion of time spent by a given person looking at each of the other people during the predetermined period, and storing the data in the database so that it is associated with the corresponding image data, sound data, text data and viewing data.

35. (Original): A method according to claim 34, wherein the predetermined period comprises a period during which the given person was talking.



36. (Original): A method according to claim 19, further comprising the step of generating a signal conveying information defining the image data selected in the camera selection step.

37. (Original): A method according to claim 31, further comprising the step of generating a signal conveying the database with data therein.

38. (Original): A method according to claim 37, further comprising the step of recording the signal either directly or indirectly to generate a recording thereof.

39.-53. (Canceled).